



# Assessment of Parameter Uncertainty in Plant Growth Model Identification

Yuting Chen, Paul-Henry Cournède

## ► To cite this version:

Yuting Chen, Paul-Henry Cournède. Assessment of Parameter Uncertainty in Plant Growth Model Identification. 2012 IEEE 4th International Symposium on Plant Growth Modeling, Simulation, Visualization and Applications, Oct 2012, China. CHEN, page 43. hal-00776551

**HAL Id: hal-00776551**

**<https://hal.science/hal-00776551>**

Submitted on 15 Jan 2013

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Assessment of Parameter Uncertainty in Plant Growth Model Identification

Yuting CHEN, Paul-Henry COURNEDE

*Ecole Centrale Paris*

*Laboratory of Applied Mathematics*

*92295 Châtenay-Malabry, France*

*Email: yuting.chen@ecp.fr*

**Abstract**—For the parametric identification of plant growth models, we generally face limited or uneven experimental data, and complex nonlinear dynamics. Both aspects make model parametrization and uncertainty analysis a difficult task. The Generalized Least Squares (GLS) estimator is often used since it can provide estimations rather rapidly with an appropriate goodness-of-fit. However, the confidence intervals are generally calculated based on linear approximations which make the uncertainty evaluation unreliable in the case of strong nonlinearity. A Bayesian approach, the Convolution Particle Filtering (CPF), can thus be applied to estimate the unknown parameters along with the hidden states. In this case, the posterior distribution obtained can be used to evaluate the uncertainty of the estimates. In order to improve its performance especially with stochastic models and in the case of rare or irregular experimental data, a conditional iterative version of the Convolution Particle Filtering (ICPF) is proposed. When applied to the Log Normal Allocation and Senescence model (LNAS) with sugar beet data, the two CPF related approaches showed better performance compared to the GLS method. The ICPF approach provided the most reliable estimations. Meanwhile, two sources of the estimation uncertainty were identified: the variance generated by the stochastic nature of the algorithm (relatively small for the ICPF approach) and the residual variance partly due to the noise models.

**Keywords**—Uncertainty analysis; Parameter estimation; Particle filtering; Plant growth model; LNAS.

## I. INTRODUCTION

Accurate parametrization of plant growth models is known to be a critical issue particularly when these models are used as predictive tools [1]. Likewise, the use of genotype specific parameters in models can only be of interest if we are able to determine significantly different parameter values for different genotypes [2]. However, most parameters cannot be measured directly or experimental protocols are heavy to implement [3] which indicate delicate parametrization. Consequently, some parameters have to be estimated from experimental data by model inversion. Such parameter estimation for plant growth models is a complex process owing to

a number of their characteristics: nonlinear dynamics, potentially with a large number of parameters, limited and irregular experimental observations and uncertain inputs (such as environmental data) [4]. Moreover, the difficulty lies not only in the point estimation of multiple unknown parameters of a complex dynamic model, but also in the evaluation of the accuracy of the estimated parameters. A confidence interval is commonly sought with the estimation value to provide some additional information concerning the reliability of the estimates. However, because of the nonlinearity and the irregular or rare measurements, although efforts have been made, there exist no general theory and methods that are capable of overcoming these difficulties properly.

In this context, the objective of this paper is to compare the properties and performance of three parameter estimation methods along with their evaluation of estimation accuracy, when applied to a typical nonlinear plant model of which the parametric identification is achieved from rare experimental data. The features of these three methods were compared with the aim of giving some new directions on the uncertainty estimation of unknown parameters for plant growth model. The three methods investigated in this paper are:

- 1) the classical Generalized Least Squares estimator (GLS) broadly used in plant growth modelling [5], with the 2-stage Aitken estimator [6], which provides point estimation and an approximation of the covariance matrix of parameter estimates;
- 2) a Bayesian approach, the Convolution Particle Filtering (CPF) [7], which appears to outperform many other classical particle filtering methods [8]. The interest of Bayesian approaches is that instead of point estimation, they provide posterior distribution for each unknown parameter, which can be simplified to credibility interval characterizing the accuracy of estimation;
- 3) the Iterative Convolution Particle Filtering (ICPF) [9], which is a new method deriving from a Bayesian framework, giving not only a point estimation of both parameters and hidden states,

but also an estimation of modelling noises. Consequently, the bootstrap method [10] can be used to evaluate the standard errors and confidence intervals of the estimates.

In Section 2, we present in detail the three parameter estimation methods. Then in section 3, the description of a plant growth model of sugar beet is given, and the results of model parameters estimated by the three methods based on real measurements are presented. The evaluation of the uncertainty of estimated parameters is conducted. The results are finally discussed, and perspectives are given towards a unifying framework for parameter estimation of plant growth models.

## II. METHODS

Plant growth models or crop models are generally written in a state-space form with deterministic dynamics. In more rare cases, modelling and measurement noises are introduced to build a stochastic model. In both situations, several estimation approaches can be used. In this paper, three suitable ones are applied.

### A. State-space models

With the objective of parametric identification, it is convenient to describe plant growth models as generalized discrete state-space models, with a state function and an observation function:

$$\begin{cases} X(t+1) = f_{t+1}(X(t), \Theta) \\ Y(t) = g_t(X(t), \Theta) \end{cases} \quad (1)$$

$X(t)$  represents the state variables at time  $t$ ,  $\Theta$  is a vector of parameters (of dimension  $p$ ) and  $Y(t)$  is the outcome vector that can be observed experimentally and usually differs from  $X(t)$  (for example plant organ masses can be observed experimentally while the daily biomass production cannot).

As presented in the following paper, some estimation methods can take advantage of a stochastic framework describing model imperfections as noises (corresponding to different sources of uncertainty). Modelling noises are represented by the series of random variables  $(\eta(t))_t$  and measurement noises by  $(\xi(t))_t$ . Generally the variables are considered as independent and identically distributed.

$$\begin{cases} X(t+1) = f_{t+1}(X(t), \Theta, \eta(t)) \\ Y(t) = g_t(X(t), \Theta, \xi(t)) \end{cases} \quad (2)$$

For biological systems, experimental observations are usually difficult to obtain, hence the system is only observed at irregular times. Let  $(t_1, t_2, \dots, t_N)$  be the  $N$  recording times. In the following paper, we denote:  $X_n = X(t_n)$  and  $Y_n = Y(t_n)$  for all  $n \in [1; N]$ . We also denote  $Z = (Y(t_1), Y(t_2), \dots, Y(t_N))$  the full observation vector of full dimension  $m$ , given by all the experimental data.

### B. Generalized Least Squares Estimator

Generalized Least Squares estimation (GLS) is a classical method traditionally used when the measurement errors have unequal variance or are correlated. Traditionally, the dynamics of the model is not taken into account in the error model and only measurement errors are considered. If we denote  $\tilde{Z}(\Theta)$  the full outcome vector of a deterministic model with parameter vector  $\Theta$  (as in the form (1)), and  $\epsilon$  the measurement error with additive assumption, we assume:

$$Z = \tilde{Z}(\Theta) + \epsilon$$

If  $\text{Var}(Z|\Theta) = \Sigma$  is known, the GLS estimator is given by:

$$\hat{\Theta}_{GLS} = \underset{\Theta}{\text{argmin}} \left( Z - \tilde{Z}(\Theta) \right)^T \Sigma^{-1} \left( Z - \tilde{Z}(\Theta) \right).$$

If the model is linear,  $\tilde{Z}(\Theta) = A\Theta$  with  $A$  a  $m \times p$  matrix, we can deduce (see for example [11]) that  $\hat{\Theta}_{GLS}$  is a Gaussian vector with variance:

$$\text{Var}(\hat{\Theta}_{GLS}) = (A^T \Sigma^{-1} A)^{-1}.$$

If the model is moderately nonlinear (see for example [12] which gives some ways to characterize the nonlinearity of plant growth models), an approximation is given by:

$$\text{Var}(\hat{\Theta}_{GLS}) = \left( \frac{\partial \tilde{Z}}{\partial \Theta} (\hat{\Theta}_{GLS})^t \Sigma^{-1} \frac{\partial \tilde{Z}}{\partial \Theta} (\hat{\Theta}_{GLS}) \right)^{-1}. \quad (3)$$

When  $\Sigma$  is unknown, the 2-stage Aitken estimator [13] proposes a way to solve the problem. If  $Z$  can be gathered into  $q$  groups, each of the error terms of in group  $Z_i$  ( $1 \leq i \leq q$ ) has common unknown variance  $\Theta_i$ ,  $\Sigma$  is thus assumed to be given by a diagonal matrix:

$$\Sigma = \begin{pmatrix} \alpha_1 I_{Z_1} & 0 & 0 & \dots & 0 \\ 0 & \alpha_2 I_{Z_2} & 0 & \dots & 0 \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ 0 & \dots & 0 & \alpha_{q-1} I_{Z_{q-1}} & 0 \\ 0 & \dots & 0 & 0 & \alpha_q I_{Z_q} \end{pmatrix}$$

with  $I_k$ , the identity matrix of rank  $k$ . This method may induce the rearrangement of the data in the  $\tilde{Z}$  vector by grouping the data of the same type.

From an algorithmic point of view, there are two stages in the estimation process:

(1)  $\alpha_i$  is estimated in the first place as the variance of all experimental data in group  $i$  to provide the first estimate  $\hat{\Sigma}^{(1)}$  of  $\Sigma$ . We obtain:

$$\hat{\Theta}_{2SA}^{(1)} = \underset{\Theta}{\text{argmin}} \left( Z - \tilde{Z}(\Theta) \right)^T \left( \hat{\Sigma}^{(1)} \right)^{-1} \left( Z - \tilde{Z}(\Theta) \right).$$

(2) We then estimate  $\alpha_i$  with:

$$\hat{\alpha}_i = \frac{1}{Z_i - p} \left( Z_i - \tilde{Z}_i \left( \hat{\Theta}_{2SA}^{(1)} \right) \right)^T \left( Z_i - \tilde{Z}_i \left( \hat{\Theta}_{2SA}^{(1)} \right) \right),$$

to obtain  $\hat{\Sigma}^{(2)}$  and the final estimator is given by

$$\hat{\Theta}_{2SA} = \operatorname{argmin}_{\Theta} \left( Z - \tilde{Z}(\Theta) \right)^T \left( \hat{\Sigma}^{(2)} \right)^{-1} \left( Z - \tilde{Z}(\Theta) \right).$$

Finally the variance of  $\hat{P}_{2SA}$  is approximated with Equation (3) using  $\hat{\Sigma}^{(2)}$ .

### C. Convolution Particle Filter for Bayesian Parameter Estimation

Let us consider the parameter estimation problem in a dynamic state-space form which can be described as a hidden Markov model. We define an augmented state vector  $X_n^a = (X_n, \Theta)$ , including  $X_n$  the hidden state at time  $n$  and  $\Theta$  the vector of unknown parameters. In the following part, if  $X$  represents a random variable with values in  $\mathcal{X}$ , then for all  $x \in \mathcal{X}$ ,  $p(x)$  will denote the probability density of  $X$  in  $x$ .

If  $Y_n$  represents the observation vector at time  $t_n$ , the first-order hidden Markov model is characterized by the transition density  $p(x_n^a | x_{n-1}^a)$  (corresponding to the state equation integrated between  $t_{n-1}$  and  $t_n$ , see [14]),  $p(y_n | x_n^a)$  (corresponding to the measurement equation) and the initial density  $p(x_0^a)$ .

Inspired by the post-regularized particle filter [15], the convolution particle filter allows the joint estimation of the parameters and the hidden states of the dynamic system, stochastic or not, from online data. The particles are initiated from an informative prior (drawn from  $p(x_0^a)$ ) or non-informative prior (distributed uniformly in the considered space). The filtering is performed recurrently, with two stages at each time step  $n$  [8]:

- **Prediction:** As an *a priori* form of the estimation, the aim is to provide a kernel estimator of  $p(x_{n+1}^a, y_{n+1} | y_{0:n})$  denoted  $\hat{p}(x_{n+1}^a, y_{n+1} | y_{0:n})$ .  $M$  particles  $\{\tilde{x}_n^{a(i)}, i = 1, \dots, M\}$  are sampled from the conditional law  $\hat{p}(x_n^a | y_{0:n})$ . We propagate each of these particles through the state equation to obtain  $\{\tilde{x}_{n+1}^{a(i)}, i = 1, \dots, M\}$  sampled from the conditional law  $p(x_{n+1}^a | \tilde{x}_n^{a(i)})$ . Thanks to  $K_{h_M^X}^X$ , the Parzen-Rosenblatt kernel with bandwidth parameter  $h_M^X$ , we are able to deduce the empirical kernel approximation of the probability density of  $(X_{n+1}^a, Y_{n+1})$  conditional to  $Y_{0:n}$ :

$$\begin{aligned} \hat{p}(x_{n+1}^a, y_{n+1} | y_{0:n}) &= \frac{1}{M} \sum_{i=1}^M K_{h_M^X}^X \left( x_{n+1}^a - \tilde{x}_{n+1}^{a(i)} \right) \\ &\quad \cdot p \left( y_{n+1} | \tilde{x}_{n+1}^{a(i)} \right). \end{aligned} \quad (4)$$

- **Correction:** This step provides an *a posteriori* form of the estimation, the kernel approximation for

$p(x_{n+1}^a | y_{1:n+1})$  is given by:

$$\begin{aligned} \hat{p}(x_{n+1}^a | y_{1:n+1}) &= \frac{1}{\sum_{i=1}^M p(y_{n+1} | \tilde{x}_{n+1}^{a(i)})} \\ &\quad \cdot \sum_{i=1}^M K_{h_M^X}^X (x_{n+1}^a - \tilde{x}_{n+1}^{a(i)}) p(y_{n+1} | \tilde{x}_{n+1}^{a(i)}). \end{aligned} \quad (5)$$

The part  $p(y_{n+1} | \tilde{x}_{n+1}^{a(i)}) / \sum_{i=1}^M p(y_{n+1} | \tilde{x}_{n+1}^{a(i)})$  can be regarded as the normalized weight  $\tilde{w}_{n+1}^{(i)}$  associated to the particle  $\tilde{x}_{n+1}^{a(i)}$ .

It is important to note that this method can be adapted in the case of rare measurements without difficulties. However, in the case that the analytic form of the observation density  $p(y_n | x_n)$  is unknown, an observation kernel can be introduced [8].

### D. Iterative Convolution Particle Filter

In the case of off-line estimation with a finite number of observations, we propose a method based on the iterative version of convolution particle filtering, which can be interpreted as an alternative of the smoothing methods [16]. It can provide the dynamic reconstruction of the model including complete observations and hidden states. After the filtering process, a set of selected particles  $\{\tilde{x}_N^{a(i)}, i = 1, \dots, M\}$  is available with their associated normalized weights  $\{\tilde{w}_N^{(i)}, i = 1, \dots, M\}$ . We repeat the filtering process by first reinitializing the particles  $\{\tilde{x}_0^{a(i)}, i = 1, \dots, M\}$ . The state vectors  $\{\tilde{x}_0^{(i)}, i = 1, \dots, M\}$  are selected in the same way as for the classical filtering process, which means that they are drawn from the distribution  $p(x_0)$ . As for the vectors of unknown parameters  $\{\tilde{\Theta}_0^{(i)}, i = 1, \dots, M\}$ , they are initialized by taking into account the estimation results of the former iteration: the posterior distribution built by  $\{\tilde{\Theta}_N^{(i)}, i = 1, \dots, M\}$  and the associated weight vector  $\{\tilde{w}_N^{(i)}, i = 1, \dots, M\}$ .

Attentions should be paid when a large part of the particles move away from the interesting region: in this case, their weights will decrease which in turn will cause degeneracy. For this reason, a threshold is selected beforehand for the effective sample size (ESS). When the ESS decreases below the threshold, a resampling procedure will be launched in order to eliminate particles with small weights and to make sure that most of the particles are close to the real trajectory [17]. The new set of particles is drawn from a multinomial distribution according to the weight vector.

After the  $l$ th filtering iteration, the parameters' and the hidden states' estimators are:

$$\hat{\Theta}^{(l)} = \sum_{i=1}^M \tilde{w}_N^{(i)} \tilde{\Theta}_N^{(i)} \quad \text{and} \quad \hat{x}_n^{(l)} = \sum_{i=1}^M \tilde{w}_n^{(i)} \tilde{x}_n^{(i)}. \quad (6)$$

Since the estimation method is defined as stochastic, after a burn-in period of  $K$  iterations, averaged estimators are chosen to take into account the fluctuations due to the stochastic algorithm [18]. Considering that the number of particles is constant, for  $l > K$ :

$$\bar{\Theta}^{(l)} = \frac{1}{l-K} \sum_{j=K+1}^l \hat{\Theta}^{(j)} \quad \text{and} \quad \bar{\hat{x}}_n^{(l)} = \frac{1}{l-K} \sum_{j=K+1}^l \hat{x}_n^{(j)}. \quad (7)$$

When the hidden Markov model is written in the form of a stochastic system, the vector  $\Theta$  consists of  $(\Theta_1, \Theta_2)$ , with  $\Theta_1$  the parameters that appear in the deterministic part of the model (state equation and measurement equation) and  $\Theta_2$  the parameters of the noise model (parameters of the stochastic distributions of  $\eta$  and  $\xi$  in (2)). In this paper, we propose a conditional version of the ICPF algorithm: in the first place, the estimation of the hidden states and of  $\Theta_1$  is performed by considering that  $\Theta_2$  is known. In practice, small noise parameters, although different from their real values, should ensure the convergence of this algorithm towards satisfactory estimation results (estimation of parameters and hidden states from the deterministic part of the model). Under this assumption, the initialization of the algorithm is therefore carried out by choosing small values for  $\Theta_2$  which stand for small noises (standard error 0.02 in our tests). Once  $\Theta_1$  is estimated as well as the hidden states, we are able to estimate directly the modelling noises and the measurement noises  $\Theta_2$ . Afterwards  $\Theta_1$  is estimated again taking into account the new computed value of  $\Theta_2$ . In this way, the algorithm can be iterated until convergence (generally 3 alternate estimations of  $\Theta_1$  and  $\Theta_2$  are sufficient).

We highlight that the successive iterations of filtering make the posterior distributions of the parameters no longer representative of the parameters' uncertainty for the remaining uncertainty is linked to the stochastic algorithm. Nevertheless, in order to obtain the related confidence intervals, parametric bootstrap method can be implemented [10]. Since the conditional version of the algorithm appears to provide fair estimates of the parameters  $\Theta_2$  ([9]), new observation vectors are randomly generated with  $\hat{\Theta}$  and the ICPF estimation is performed for each new data set. This method allows us to evaluate the uncertainty related to the estimation  $\hat{\Theta}$ . Likewise, since particle filtering methods are stochastic algorithms, we also applied our approach to the same experimental data set a large number of times in order to evaluate the algorithmic uncertainty.

### III. MODEL AND DATA

#### A. LNAS Model of Plant Growth

In this section, a Log Normal Allocation and Senescence (LNAS) daily crop model is introduced in its

stochastic version (with modelling and measurement errors). Inspired by the model presented in [19], the equations are specifically derived for the sugar beet, per unit surface area, with two kinds of organ compartments: foliage and root system. Note that the equations can be adapted to other type of plant without difficulty by adding other types of compartments.

**Biomass production:**  $Q(t)$  is the biomass production on day  $t$  per unit surface area ( $g.m^{-2}$ ) obtained by generalizing the Beer-Lambert law [20]: the fraction of intercepted radiation is given by  $(1 - e^{-\lambda \cdot Q_g(t)})$ , with  $\lambda$  ( $g^{-1}.m^2$ ) a known parameter and  $Q_g(t)$  the total mass of green leaves on day  $t$  (in  $g.m^{-2}$ ). The biomass production of the whole plant is then deduced by multiplying the total amount of absorbed photosynthetically active radiation per unit surface area (PAR, in  $MJ.m^{-2}$ ) and an energetic efficiency  $\mu$  (in  $g \cdot MJ^{-1}$ ):

$$Q(t) = \left( \mu \cdot PAR(t) \left( 1 - e^{-\lambda Q_g(t)} \right) \right) \cdot (1 + \eta_Q(t)) \quad (8)$$

with the modelling noise  $\eta_Q \sim \mathcal{N}(0, \sigma_Q^2)$ .

**Allocation** between the foliage and root system compartments:

$$Q_f(t+1) = Q_f(t) + \gamma(t) \cdot Q(t) \quad (9)$$

$$Q_r(t+1) = Q_r(t) + (1 - \gamma(t)) \cdot Q(t) \quad (10)$$

The function  $\gamma$  is defined as:

$$\gamma(t) = (\gamma_0 + (\gamma_f - \gamma_0) \cdot G_a(\tau(t))) \cdot (1 + \eta_\gamma(t)) \quad (11)$$

with  $\tau(t)$  the thermal time (according to the accumulated daily temperature since emergence day),  $G_a$  the cumulative distribution function of a log-normal law, with an underlying normal distribution characterized by  $\mu_a$ ,  $s_a$ , and the modelling noise (process noise)  $\eta_\gamma(t) \sim \mathcal{N}(0, \sigma_\gamma^2)$ .

**Senescence:** The senescent foliage mass  $Q_s$  is a proportion of the accumulated foliage mass determined by another cumulative distribution of a log-normal law characterized by  $\mu_s$ ,  $s_s$ :

$$Q_s(t) = G_s(\tau(t) - \tau_{sen}) Q_f(t) \quad (12)$$

with  $\tau_{sen}$  the thermal time at which the senescence process initiates. The green foliage mass  $Q_g$  can be hence obtained easily:

$$Q_g(t) = Q_f(t) - Q_s(t) \quad (13)$$

**Observations:** The observation variables are potentially available from field measurements:

$$Y(t) = \begin{pmatrix} Q_g(t) \cdot (1 + \epsilon_g(t)) \\ Q_s(t) \cdot (1 + \epsilon_s(t)) \\ Q_r(t) \cdot (1 + \epsilon_r(t)) \end{pmatrix} \quad (14)$$

with measurement noises:  $\epsilon_g(t) \sim \mathcal{N}(0, \sigma_g^2)$ ,  $\epsilon_s(t) \sim \mathcal{N}(0, \sigma_s^2)$  and  $\epsilon_r(t) \sim \mathcal{N}(0, \sigma_r^2)$ .

## B. Experimental Data

The data used in this analysis were obtained by the French institute for Sugar Beet research in 2008 as presented by [21]. They are based on the measurements of 20 individual plants at days 23, 45, 52, 60, 73, 107 and 143 after plant emergence. For each plant,  $Q_g$  the green foliage mass,  $Q_s$  the senescent foliage mass and  $Q_r$  the root compartment mass were measured. The observation vector  $Y$  is obtained by averaging each data based on the 20 samples and extrapolated at  $m^2$  level by being multiplied by the observed density.

## IV. RESULTS

### A. Fitting Results

For all the three approaches, the unknown parameter vector for the deterministic part of the model was  $\Theta_1 = (\mu, \mu_a, s_a, \mu_s, s_s)$ . For the CPF initialization, 10 values were drawn from a non-informative prior distribution uniformly for each parameter which formed a grid of  $10^5$  combinations of values for  $\Theta_1$ . Each combination was assigned to 5 particles which therefore made the initial number of particles 500000 ( $= 10^5 \star 5$ ).

Regarding the conditional version of ICPF approach, the unknown noise parameter vector was  $\Theta_2 = (\sigma_Q, \sigma_\gamma, \sigma_g, \sigma_s, \sigma_r)$ . 8 values were drawn from the uniform distribution for each parameter of  $\Theta_1$  which made the initial number of the particles 32768 ( $= 8^5$ ). The procedure began with the estimation of  $\Theta_1$  given  $\Theta_2$  considering known, then  $\Theta_2$  was estimated based on the estimates of hidden states. The process was then conducted with the new value of  $\Theta_2$  and iterated. Finally, 3 repetitions of the conditional version of ICPF were implemented in our test, each of them contained 600 iterations.

Table I presents the first estimation results. In this result, each parameter estimation method was applied once to the LNAS model with the given 2008 observation vector.

| Estimation of the Parameters $\Theta_1$ |      |      |      |
|---|------|------|------|
| Parameter                               | GLS  | CPF  | ICPF |
| $\mu$                                   | 2.97 | 3.06 | 2.97 |
| $\mu_a$                                 | 689  | 671  | 695  |
| $s_a$                                   | 225  | 342  | 320  |
| $\mu_s$                                 | 3450 | 2542 | 2469 |
| $s_s$                                   | 2570 | 1141 | 969  |

Table I

*Estimates of the parameters from the deterministic part of the LNAS model by the GLS, the CPF (500000 particles) and the conditional ICPF (32768 particles, 600 iterations, 3 repetitions) approaches.*

Before comparing the fitting results of the 3 methods, we note that the two CPF related methods have also estimated the hidden states of the model while GLS hasn't been able to. Therefore, for the latter, the estimated parameters were used to simulate the

observations. Thus, the comparison was conducted by using estimated observations (for CPF and ICPF) and simulated observations (for GLS) confronted to the experimental data, which contains senescent foliage mass  $Q_s$ , green foliage mass  $Q_g$  and root mass  $Q_r$  (Fig. 1).

From a goodness-of-fit point of view, the three approaches provided satisfactory results and globally met the demand of the parametrization for the LNAS model, despite the different estimates given for several parameters. However, we remarked the resemblance of the estimates given by ICPF and GLS for  $\mu$  and  $\mu_a$ . Moreover, the estimation given by the GLS method showed an excellent performance regarding the fitting of  $Q_g$ , but not for  $Q_s$  which probably resulted from a poor estimation of  $\mu_s$  and  $s_s$ . The CPF and the ICPF gave different estimates for all the parameters, nevertheless, their estimation for the hidden states showed great similarity. This result suggests that several combinations of parameters might lead to comparable outcome.

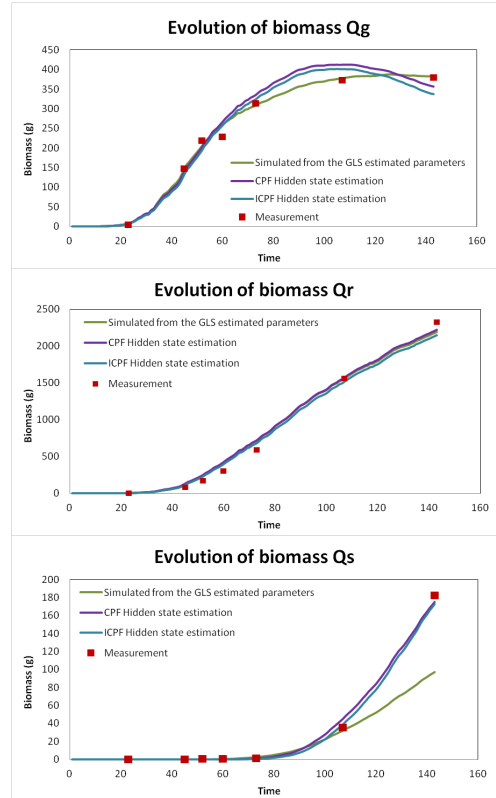


Figure 1. *Comparison of experimental data and estimated data of green foliage mass  $Q_g$ , root mass  $Q_r$ , and senescent foliage mass  $Q_s$  with the 3 methods: GLS, the CPF (500000 particles) and the conditional ICPF (32768 particles, 600 iterations, 3 repetitions) approaches.*

### B. Algorithmic Uncertainty for Particle Methods

For the purpose of evaluating the bias due to the stochastic algorithm of the CPF approach, the estimation algorithm were performed 100 times with the same

| CPF<br>$M = 32768 (8^5)$    | Single test |                       | 100 Repetitions |                       |
|-----------------------------|-------------|-----------------------|-----------------|-----------------------|
|                             | Mean        | 95%CI*                | Mean            | 95%CI                 |
| $\mu$                       | 3.04        | [ 2.96 ; 3.11 [       | 3.05            | [ 3.03 ; 3.07 [       |
| $\mu_a$                     | 667.12      | [ 650.09 ; 685.07 [   | 668.98          | [ 665.98 ; 672.57 [   |
| $s_a$                       | 342.03      | [ 287.67 ; 393.06 [   | 342.53          | [ 340.54 ; 344.76 [   |
| $\mu_s$                     | 2431.30     | [ 2666.30 ; 2545.34 [ | 2542.61         | [ 2538.80 ; 2545.90 [ |
| $s_s$                       | 1159.65     | [ 994.82 ; 1311.80 [  | 1150.67         | [ 1142.20 ; 1158.00 [ |
| $M = 100000 (10^5)$         | Mean        | 95%CI*                | Mean            | 95%CI                 |
| $\mu$                       | 3.04        | [ 2.97 ; 3.11 [       | 3.05            | [ 3.03 ; 3.07 [       |
| $\mu_a$                     | 667.12      | [ 651.30 ; 683.61 [   | 669.08          | [ 666.15 ; 671.96 [   |
| $s_a$                       | 342.95      | [ 289.19 ; 392.57 [   | 342.47          | [ 340.95 ; 344.14 [   |
| $\mu_s$                     | 2541.53     | [ 2435.60 ; 2653.20 [ | 2542.74         | [ 2540.90 ; 2544.80 [ |
| $s_s$                       | 1147.37     | [ 993.38 ; 1306.20 [  | 1148.27         | [ 1144.10 ; 1152.50 [ |
| $M = 500000 (10^5 \star 5)$ | Mean        | 95%CI*                | Mean            | 95%CI                 |
| $\mu$                       | 3.04        | [ 2.99 ; 3.10 [       | 3.05            | [ 3.03 ; 3.06 [       |
| $\mu_a$                     | 668.57      | [ 655.17 ; 682.43 [   | 669.11          | [ 666.49 ; 672.40 [   |
| $s_a$                       | 343.67      | [ 289.22 ; 391.68 [   | 342.52          | [ 340.99 ; 344.12 [   |
| $\mu_s$                     | 2543.36     | [ 2445.10 ; 2647.50 [ | 2542.87         | [ 2541.10 ; 2544.60 [ |
| $s_s$                       | 1142.14     | [ 994.77 ; 1304.50 [  | 1147.90         | [ 1142.60 ; 1151.70 [ |
| <hr/>                       |             |                       |                 |                       |
| ICPF (3 repetitions)        |             | Bootstrap (100 tests) |                 | 100 Repetitions       |
| $M = 32768 (8^5)$           | Mean        | 95%CI                 | Mean            | 95%CI                 |
| $\mu$                       | 2.94        | [ 2.69 ; 3.17 [       | 2.97            | [ 2.96 ; 2.97 [       |
| $\mu_a$                     | 710.01      | [ 611.70 ; 822.08 [   | 694.58          | [ 692.52 ; 696.51 [   |
| $s_a$                       | 322.62      | [ 165.62 ; 483.79 [   | 319.73          | [ 315.40 ; 324.06 [   |
| $\mu_s$                     | 2499.09     | [ 2319.30 ; 2713.90 [ | 2476.71         | [ 2470.90 ; 2482.00 [ |
| $s_s$                       | 1025.67     | [ 763.95 ; 1349.60 [  | 985.64          | [ 974.85 ; 996.20 [   |
|                             | Mean        | Std.                  | Mean            | Std.                  |
| $\text{std}(\eta_Q)$        | 0.0204      | 0.0018                | 0.0198          | 0.0023                |
| $\text{std}(\eta_\gamma)$   | 0.0207      | 0.0019                | 0.0207          | 0.0019                |
| $\text{std}(\epsilon_g)$    | 0.0821      | 0.0300                | 0.1071          | 0.0038                |
| $\text{std}(\epsilon_s)$    | 0.0374      | 0.0277                | 0.0649          | 0.0030                |
| $\text{std}(\epsilon_r)$    | 0.1797      | 0.0857                | 0.1632          | 0.0037                |

Table II

Estimates of the parameters from the deterministic part of the LNAS model provided by the CPF and the conditional ICPF approaches running with different number of particles  $M$ . The 100 repetitions were performed with the same observation data. For the CPF method, The 95% Credible Interval (CI)  $\star$  was obtained by the posterior distribution. For the ICPF, the bootstrap confidence interval was based on 100 estimations from virtual observations generated with the stochastic model parametrized with the originally estimated parameters.

observation data. Thus, the 95% confidence intervals provided for each average estimates may present the uncertainty purely resulting from the stochasticity of the method (such algorithmic uncertainty does not exist for GLS).

Meanwhile, the effect of using different numbers of particles was tested which may help us to understand its influence over the estimation performance of CPF (Table II). The credible intervals based on the posterior distributions obtained by the list of particles are also presented in Table II. According to the confidence intervals based on the 100 repetitions, we highlight that the variance related to the stochastic nature of the method (algorithmic variance) remains quite small. When the number of particles increased from 32768 to 500000, the credible intervals as well as the confidence intervals shrank slightly for CPF. Nevertheless, it seems to reach a limit when the number of particles increases (eg. CI of  $\mu_a$  and  $s_s$  based on 100 repetitions). On the other hand, not only the estimates given by the three single tests were relatively close, but the mean estimates based on 100 repetitions were almost the same in the three different situations. This result may suggest that the augmentation of the particle number can improve the

estimation performance to a certain level, but afterwards it might only become a burden of calculation.

The estimates provided by the bootstrap test of the conditional ICPF approach were distinguished from the others (Table II), since it took into account both modelling and measurement noises during the estimation process. The 100 repetitions of the same conditional ICPF test showed the part of variance due to the method. The remaining part of the variance in the bootstrap confidence interval could be explained by the part of the information that the model failed to explain or the lack of data.

### C. Final Estimation Results

Table III illustrates the final estimation results of the three methods. The point estimations were given with their level of uncertainty characterized by 95% credible interval (for CPF) or confidence interval (for ICPF, GLS). For all the parameters, CPF gave the narrowest CI. ICPF had the largest intervals for  $\mu$ ,  $\mu_a$ ,  $s_a$ , while for  $\mu_s$  and  $s_s$ , GLS provided very high estimation uncertainty, which suggests a non-reliable estimation of the parameters (as also illustrated by the fitting results of  $Q_s$  in Fig. 1 ).

|         | Bootstrap test of ICPF |                     | single CPF (500000 particles) |                       | GLS  |                 |
|---------|------------------------|---------------------|-------------------------------|-----------------------|------|-----------------|
|         | Mean                   | 95%CI               | Mean                          | 95%CI                 | Mean | 95%CI           |
| $\mu$   | 2.97                   | [2.69 ; 3.17 [      | 3.05                          | [ 2.99 ; 3.10 [       | 2.97 | [ 2.86 ; 3.08 ] |
| $\mu_a$ | 694.58                 | [611.70 ; 822.08[   | 669.11                        | [ 655.17 ; 682.43 [   | 689  | [650 ; 728]     |
| $s_a$   | 319.73                 | [165.62 ; 483.79 [  | 342.52                        | [ 289.22 ; 391.68 [   | 225  | [81 ; 370]      |
| $\mu_s$ | 2476.71                | [2319.30 ; 2713.90[ | 2542.87                       | [ 2445.10 ; 2647.50 [ | 3450 | [1450;5450]     |
| $s_s$   | 985.64                 | [763.95 ; 1349.60[  | 1147.90                       | [ 994.77 ; 1304.50 [  | 2570 | [0;9370] *      |

Table III

*Estimates of the deterministic part parameters of LNAS model obtained by the conditional ICPF (32768 particles, 600 iterations, 3 repetitions), the CPF ( $5 \times 10^6$  particles) and the GLS approaches. The mean values and the Confidence Intervals (CI) provided by the conditional ICPF approach were based on 100 bootstrap tests. The estimates given by GLS method were obtained with the GLS Estimator (2-stage Aitken estimator), which failed to provide a proper confidence bound of  $s_s$  (\*:).*

It deserves to note that the confidence intervals given by ICPF also contain the parameters estimated by the two other methods (except the non-reliable senescence parameters of GLS), while reversely, the mean estimates of CPF are the only ones that are contained in all the credible / confidence intervals given by the other methods.

#### D. Robustness Test

In this paragraph, an interesting result is presented to compare the robustness of the particle filtering methods. The results of the parameter estimation from real experimental data were used to generate virtual observation data sets. Two particle filtering methods were tested with these virtual data sets. For the sake of simplicity, we do not consider the GLS method due to its former poor performance. ICPF was performed with virtual data generated with parameters previously estimated by CPF, and conversely CPF was performed with virtual data generated with ICPF parameter estimations. Each of these two tests was repeated 100 times based on the same original 2008 data set. The mean estimates and confidence intervals are given in Table IV. It is meaningful to note that from this perspective, ICPF appears to be far more reliable than CPF. Although provided narrower uncertainty intervals, for parameter  $\mu$ ,  $\mu_a$  and  $s_s$  CPF failed to supply CIs which contain their real values used in the generation of the data sets! On the contrary, ICPF estimates were very close to their real values, which were as well contained in the estimated confidence intervals.

#### V. DISCUSSION AND CONCLUSION

All the three proposed methods were able to estimate the parameters of the LNAS model from a limited number of measurements and with satisfactory goodness-of-fit. Yet not all the estimation results were appropriate, especially the GLS method that failed to provide reliable estimate for the senescence process.

From the estimation results based on the 2008 experimental data, it appears to be complicated to choose the best method between CPF and ICPF. A quick look

would lead us to consider the CPF approach since it has provided the narrowest uncertainty interval and with estimates belonging to the confidence intervals computed by all the other methods. However, the robustness test from virtual data in section IV-D contradicts this assumption as the robustness of the method seemed to be quite poor and the credible intervals given were proved to be not reliable. Evidently, such conclusion could be reconsidered in a situation with richer data set.

Finally, the ICPF method appears quite reliable both in terms of its point estimation and confidence intervals, as well as its hidden states estimations. The relatively considerable estimated measurement noises are probably due to the residual part of the information that the model failed to explain.

Meanwhile, it is meaningful to remark the importance of the noise model for the ICPF method, particularly when applied to LNAS model, the multiplicative assumption of the noise models is worth discussing: when the biomass increases, this assumption indicates large noises which may twist the estimations.

Furthermore, regarding the computation time, the conditional ICPF is clearly the most time-consuming approach (18 hours for performing a conditional ICPF test with 32768 particles and 3 repetitions of 600 iterations on a standard computer), which is the most important shortcoming of this method. The CPF and the GLS methods, on the other hand, are much more efficient. An intriguing approach could be the combination of these two methods: using GLS in the first place could help to identify the zone of interest for the unknown parameters in order to reduce the number of particles necessary.

Finally, uncertainty analysis and tests of the predictive capacity of the parametrized model would be the next step for a full evaluation of the estimation method.

#### REFERENCE

- [1] C. Spitters, "Crop growth models: Their usefulness and limitations," *Acta Horticulturae*, vol. 267, pp. 349–368, 1990.



|         | Estimation with ICPF |         |                     | Estimation with CPF |         |                       |
|---------|----------------------|---------|---------------------|---------------------|---------|-----------------------|
|         | Real Value           | Mean    | 95%CI               | Real Value          | Mean    | 95%CI                 |
| $\mu$   | 3.05                 | 3.06    | [2.82 ; 3.31[       | 2.97                | 3.03    | [ 2.98 ; 3.10 [       |
| $\mu_a$ | 669.11               | 670.46  | [572.12 ; 768.79[   | 694.58              | 666.48  | [ 656.04 ; 679.54 [   |
| $s_a$   | 342.52               | 341.66  | [180.20 ; 503.12 [  | 319.73              | 339.97  | [ 309.80 ; 369.61 [   |
| $\mu_s$ | 2542.87              | 2544.96 | [2360.41 ; 2729.51[ | 2476.71             | 2534.29 | [ 2454.20 ; 2624.00 [ |
| $s_s$   | 1147.90              | 1149.60 | [820.49 ; 1478.71[  | 985.64              | 1134.85 | [ 1160.40 ; 1184.50 [ |

Table IV

*LNAS parameter estimation from virtual data: we estimated with ICPF from virtual data sets generated with the parameters provided by CPF based on the 2008 experimental data, and conversely we estimated with CPF from virtual data sets generated with the parameters provided by ICPF based on the same 2008 experimental data.*

- [2] M.-H. Jeuffroy, A. Barbottin, J. Jones, and J. Lecoecur, "Crop models with genotype parameters," in *Working with Dynamic Crop Models*, 2006.
- [3] J. Lecoecur, R. Poiré-Lassus, A. Christophe, B. Palas, P. Casadebaig, P. Debaeke, F. Vear, and L. Guilioni, "Quantifying physiological determinants of genetic variation for yield potential in sunflower. sunflo: a model-based analysis," *Functional Plant Biology*, vol. 38, pp. 246–259, 2011.
- [4] D. Makowski, J. Hillier, D. Wallach, B. Andrieu, and M.-H. Jeuffroy, "Parameter estimation for crop models," in *Working with Dynamic Crop Models*, 2006.
- [5] Y. Guo, Y. Ma, Z. Zhan, B. Li, M. Dingkuhn, D. Luquet, and P. de Reffye, "Parameter optimization and field validation of the functional-structural model greenlab for maize," *Annals of Botany*, vol. 97, pp. 217–230, 2006.
- [6] P.-H. Cournède, V. Letort, A. Mathieu, M.-Z. Kang, S. Lemaire, S. Trevezas, F. Houllier, and P. de Reffye, "Some parameter estimation issues in functional-structural plant modelling," *Mathematical Modelling of Natural Phenomena*, vol. 6, no. 2, pp. 133–159, 2011.
- [7] V. Rossi, *Filtrage non linéaire par noyaux de convolution - Application à un procédé de dépollution biologique*. PhD thesis, Ecole Nationale Supérieure Agronomique de Montpellier, 2004.
- [8] F. Campillo and V. Rossi, "Convolution Particle Filter for Parameter Estimation in General State-Space Models," *IEEE Transactions in Aerospace and Electronics*, vol. 45, no. 3, pp. 1063–1072, 2009.
- [9] Y. Chen, B. Bayol, C. Loi, S. Trevezas, and P.-H. Cournède, "Filtrage par noyaux de convolution itératif," in *Actes des 44èmes Journées de Statistique JDS2012, Bruxelles 21-25 Mai 2012*, 2012.
- [10] B. Efron and R. Tibshirani, *An Introduction to the Bootstrap*. Chapman & Hall/CRC Monographs on Statistics and Applied Probability, 1994.
- [11] C. Rao, *Linear Statistical Inference and Its Applications*. Wiley, New York, 1973.
- [12] Q. Wu, P.-H. Cournède, and A. Mathieu, "An efficient computational method for global sensitivity analysis and its application to tree growth modelling," *Reliability Engineering & System Safety*, vol. In Press:DOI: 10.1016/j.ress.2011.07.001, 2011.
- [13] W. Taylor, "Small sample properties of a class of two-stage Aitken estimator," *Econometrica*, vol. 45, no. 2, pp. 497–508, 1977.
- [14] M. Quach, N. Brunel, and F. d'Alché Buc, "Estimating parameters and hidden variables in nonlinear state-space models based on ODEs for biological networks inference," *Bioinformatics*, vol. 23, no. 23, pp. 3209–3216, 2007.
- [15] N. Oudjane and C. Musso, "Regularized particle schemes applied to the tracking problem," in *International Radar Symposium, Munich, Proceedings*, 1998.
- [16] G. Kitagawa, "Monte Carlo filter and smoother for non-Gaussian nonlinear state space models," *Journal of Computational and Graphical Statistics*, vol. 5, no. 1, pp. 1–25, 1996.
- [17] A. Kong, J. Liu, and W. Wong, "Sequential Imputations and Bayesian Missing Data Problems," *Journal of the American Statistical Association*, vol. 89, no. 425, pp. 278–288, 1994.
- [18] O. Cappé, E. Moulines, and T. Rydén, *Inference in Hidden Markov Models*. 2005.
- [19] S. Lemaire, F. Maupas, P.-H. Cournède, and P. de Reffye, "A morphogenetic crop model for sugar-beet (*beta vulgaris* L.)," in *International Symposium on Crop Modeling and Decision Support: ISCMDS 2008, April 19-22, 2008, Nanjing, China*, 2008.
- [20] T. Nilson, "A theoretical analysis of the frequency of gaps in plant stands," *Agricultural and Forest Meteorology*, vol. 8, pp. 25–38, 1971.
- [21] S. Lemaire, F. Maupas, P.-H. Cournède, J.-M. Allirand, P. de Reffye, and B. Ney, "Analysis of the density effects on the source-sink dynamics in sugar-beet growth," in *3rd international symposium on Plant Growth and Applications (PMA09), Beijing, China* (B.-G. Li, M. Jaeger, and Y. Guo, eds.), IEEE Computer Society (Los Alamitos, California), November 9-12 2009.